Descent Property of ISTA via Majorization-Minimization

İlker Bayram

Let y, A be given. Consider the minimization of :

$$J(x) = \frac{1}{2}\,\|y - A\,x\|_2^2 + \lambda\,\|x\|_1. \tag{1}$$

Also let the matrix A satisfy $A^T\,A < \alpha I$ for some scalar α.

Majorization-minimization idea goes as follows. Suppose we have an estimate of the minimizer of $J(\cdot)$, namely $\bar{x}$. We would like to find another point x^* such that $J(x^*) < J(\bar{x})$. Let

$$M(\bar{x}, x) = \frac{1}{2}(\bar{x} - x)^T\,(\alpha\,I - A^T\,A)\,(\bar{x} - x), \tag{2}$$

and consider the function

$$\bar{J}(x) = J(x) + M(\bar{x}, x). \tag{3}$$

Notice that

(i) $M(\bar{x}, x) \geq 0$,

(ii) $M(\bar{x}, \bar{x}) = 0$.

These two conditions imply that,

$$\left[\min_x \bar{J}(x)\right] \leq J(x). \tag{4}$$

Therefore if we set

$$x^* = \arg\min_x \bar{J}(x), \tag{5}$$

then $J(x*) \leq J(\bar{x})$ – that is, we achieved descent in the cost function. We can now apply this trick again on x^* to further decrease the cost function.

Let us now look at the minimization of $\bar{J}(x)$. First, we note that the new function $\bar{J}(x)$ is separable in its entries, i.e.,

$$\bar{J}(x) = \sum_i \frac{\alpha}{2}\,x_i^2 - c_i\,x_i + \lambda\,|x_i| + \text{const.} \qquad \text{(check this!)} \tag{6}$$

where

$$c = \alpha\,\bar{x} + A^T\,y - A^T\,A\,\bar{x} \tag{7}$$

and the term 'const.' is independent of x. We also note that the minimizer of the scalar function

$$f(x) = \frac{1}{2}\,(x - w)^2 + \gamma\,|x| \tag{8}$$

is given as

$$z = \text{soft}(w, \gamma) \tag{9}$$

where

$$\text{soft}(w, \gamma) = \text{sign}(w)\,\max(|w| - \gamma, 0). \tag{10}$$

This motivates the following descent algorithm for minimizing of $J(\cdot)$.

Algorithm 1 Iterated Shrinkage Thresholding Algorithm

Initialize x. Set $\gamma \leftarrow \lambda/\alpha$.
repeat
  $c \leftarrow x + \frac{1}{\alpha}\,A^T(y - A\,x)$.
  $x_i \leftarrow \text{soft}(c_i, \gamma)$.
until convergence